(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平10－91191

(43)公開日　平成10年(1998)４月10日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　5/04		Ｇ１０Ｌ　5/04	Ｆ

審査請求　未請求　請求項の数６　ＯＬ　（全 12 頁）

(21)出願番号　特願平8－246718

(22)出願日　平成８年(1996)９月18日

(71)出願人　000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地

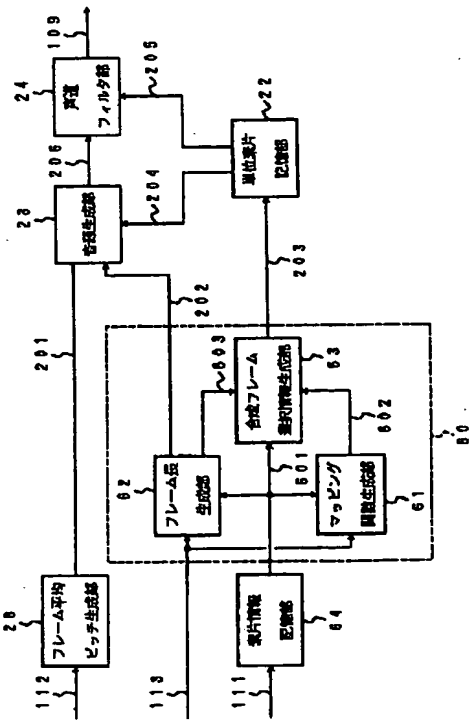
(72)発明者　篭嶋　岳彦
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内

(72)発明者　赤嶺　政巳
神奈川県川崎市幸区小向東芝町１番地　株式会社東芝研究開発センター内

(74)代理人　弁理士　鈴江　武彦　（外６名）

(54)【発明の名称】　音声合成方法

(57)【要約】
【課題】自然性を損なうことなく任意の音韻継続時間長の合成音声を生成できる音声合成方法を提供する。
【解決手段】合成しようとする音声信号の性質を表す第１のフレーム単位の合成パラメータ（音源パラメータ２０４、声道パラメータ２０５）を用いて音声信号１０９を合成する音声合成方法において、単位素片記憶部２２に合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、入力された音韻継続時間長１１３の情報に従ってマッピング関数生成部６１で生成されたマッピング関数を合成フレーム選択情報生成部６３に与えて、第１のフレームと第２のフレームとの対応付けを行い、かつマッピング関数に基づいて第１のフレームに対応する第２のフレームの単位素片を間引くかまたは繰り返すことにより、合成パラメータを生成する。



【特許請求の範囲】
【請求項１】合成しようとする音声信号の性質を表す第１のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、
前記合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、
入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第１のフレームと第２のフレームとの対応付けを行い、
第１のフレームに対応する第２のフレームの単位素片を前記合成パラメータとして生成することを特徴とする音声合成方法。
【請求項２】合成しようとする音声信号の性質を表す第１のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、
前記合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、
入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第１のフレームと第２のフレームとの対応付けを行い、
かつ該マッピング関数に基づいて第１のフレームに対応する第２のフレームの単位素片を間引くかまたは繰り返すことにより、前記合成パラメータを生成することを特徴とする音声合成方法。
【請求項３】合成しようとする音声信号の性質を表す第１のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、
前記合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、
入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第１のフレームと第２のフレームとの対応付けを行うとともに、
該マッピング関数に基づいて第１のフレームのフレーム長を変化させ、
第１のフレームに対応する第２のフレームの単位素片を前記合成パラメータとして生成することを特徴とする音声合成方法。
【請求項４】合成しようとする音声信号の性質を表す第１のフレーム単位の合成パラメータを用いて音声信号を合成する音声合成方法において、
前記合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、
入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第１のフレームと第２のフレームとの対応付けを行い、
前記単位素片の素片長が前記音韻継続時間長より小さい場合は該マッピング関数に基づいて第１のフレームに対応する第２のフレームの単位素片を間引くかまたは繰り返すことにより前記合成パラメータを生成し、
前記単位素片の素片長が前記音韻継続時間長以上の場合には該マッピング関数に基づいて第１のフレームのフレーム長を変化させ、
第１のフレームに対応する第２のフレームの単位素片を前記合成パラメータとして生成することを特徴とする音声合成方法。
【請求項５】前記マッピング関数として、その傾きが音韻境界付近で予め定められた略一定の値となる関数を用いることを特徴とする請求項１～４のいずれか１項に記載の音声合成方法。
【請求項６】前記マッピング関数を前記単位素片の過渡部と定常部を指定する情報および前記音韻継続時間長とに従って生成することを特徴とする請求項１～４のいずれか１項に記載の音声合成方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、テキスト音声合成のための音声合成方法に係り、特に音韻記号列、ピッチおよび音韻継続時間長などの情報から音声信号を生成する音声合成方法に関する。
【０００２】
【従来の技術】任意の文章から人工的に音声信号を作り出すことをテキスト音声合成という。テキスト音声合成システムは、一般的に図８に示されるように、言語処理部１、音韻処理部２および音声合成器３の３つの要素から構成される。入力されたテキスト１１５は、まず言語処理部１において形態素解析や構文解析などが行われ、次に韻律処理部２においてアクセントやイントネーションの処理が行われて、音韻記号列１１１、ピッチパターン１１２および音韻継続時間長１１３などの情報が出力される。最後に、音声合成器３で音節、音素、１ピッチ区間などの基本となる小さな単位（単位素片）の特徴パラメータを音韻記号列、ピッチパターン、音韻継続時間長などの情報に従って選択し、ピッチや継続時間長を制御して接続することによって合成音声信号１０９を生成する。
【０００３】自然音声では、発話の速度、意味内容、前後の音韻などの影響で、音韻の継続時間長が様々に変化している。そこで、自然音声に近い高品質な音声を合成するためには、任意の継続時間長の音声を合成器によって合成することが可能でなければならない。
【０００４】このような継続時間長を変化させることができる音声合成器として、特開昭６１－２９０４９９「発声速度可変の音声合成装置」が知られている。この音声合成器の原理は、音声の定常部において単位素片のフレームを間引いて合成フレーム列を構成するとともに、合成フレームのフレーム長を一様に変化させることによって継続時間長を変化させるものである。
【０００５】この従来の音声合成器の構成について図９を用いて説明する。図９において、フレーム平均ピッチ生成部２８は、ピッチパターン１１２よりフレーム平均

ピッチ周期２０１を求めて出力する。素片情報記憶部２５は、単位素片フレームのフレーム数、定常部の範囲および過渡部の範囲などの素片情報を記憶しており、入力された音韻記号列１１１に従って、対応する素片の素片情報２０７を出力する。フレーム長生成部２６は、素片情報２０７に基づいて継続時間長が音韻継続時間長１１３と等しくなるようにフレーム長２０２と定常部伸縮率２０８を求めて出力する。合成フレーム選択情報生成部２７は、素片情報２０７と定常部伸縮率２０８に従って合成フレームのフレーム選択情報２０３を出力する。

【０００６】単位素片記憶部２２は、例えば合成単位を音素とすると、全ての母音および全ての子音に対応する単位素片の情報を記憶している。各単位素片は、例えば当該音韻の音声信号を一定のフレーム周期で分析することによって生成され、複数の単位素片フレームにより構成される。各単位素片フレームは音源パラメータおよび声道パラメータから構成されており、フレーム選択情報

$$(v_{na} + n_b)\ f = T$$

合成フレーム情報生成部２７では、単位素片フレームのうち定常部のフレームについて、定常部伸縮率２０８に従って間引きまたは繰り返しを行って合成フレーム列を生成し、各フレームのフレーム選択情報２０３を出力する。このようにして、任意の音韻継続時間長の音声を合成することが可能である。

【００１０】

【発明が解決しようとする課題】上述した従来の音声合成器では、音声の定常部の間引きと過渡部の伸縮によって音韻継続時間長の変化を実現しているが、自然音声においては必ずしもこのようなモデルで音韻継続時間の異なる音声を表すことはできない。すなわち、従来の音声合成器は明瞭性を保つには有効であるが、そもそも自然音声では定常部と過渡部の境界は明確ではなく、性質が常に滑らかに変化しているため、過渡部と定常部に分類して異なる処理を行うことは、自然性が低下する原因となる。

【００１１】また、自然音声を分析することによって単位素片を生成する場合、合成単位のフレームを過渡部と定常部に分類する必要があり、このような分類は視察によって行わざるを得ないため、多大な労力を必要とする。

【００１２】本発明は、このような従来の問題点を解決するためになされたもので、自然性を損なうことなく任意の音韻継続時間長の合成音声を生成できる音声合成方法を提供することを目的とする。

【００１３】

【課題を解決するための手段】上記の課題を解決するため、本発明は合成しようとする音声信号の性質を表す第１のフレーム単位の合成パラメータを用いて音声信号を合成するに際し、合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、入力された音韻継続時間長の情報に従って生成されたマッピング関数に基づいて第１のフレームと第２のフレームとの対応付けを行い、第１のフレームに対応する第２のフレームの単位素片を合成パラメータとして生成することを特徴とする。

【００１４】ここで、「合成パラメータ」とは、声道特性、音源特性、１ピッチ長の音声波形など、音声信号の性質を表すパラメータを表している。また、子音をＣ、母音をＶでそれぞれ表すと、「単位素片」とは、Ｃ，Ｖ，ＣＶ，ＶＣＶ，ＣＶＣなどの合成単位に対応して生成された合成パラメータの系列を表すものとする。

【００１５】本発明の第１の態様では、第１のフレームと第２のフレームとの対応付けを行った後、マッピング関数に基づいて第１のフレームに対応する第２のフレームの単位素片を間引くかまたは繰り返すことにより、合成パラメータを生成する。

【００１６】第１のフレームと第２のフレームとの対応付けを行うマッピング関数は、音韻継続時間長に基づいて生成されるため、このマッピング関数に基づいて第１のフレームに対応する第２のフレームの単位素片を間引くかまたは繰り返すことで合成パラメータを生成すれば、合成音声信号の継続時間長が変化する。

【００１７】本発明の第２の態様では、第１のフレームと第２のフレームとの対応付けを行うとともに、マッピング関数に基づいて第１のフレームのフレーム長を変化させ、第１のフレームに対応する第２のフレームの単位素片を合成パラメータとして生成する。

【００１８】このようにマッピング関数に基づいて第１のフレームのフレーム長を変化させることによっても、合成音声信号の継続時間長が変化する。本発明の第３の態様では、第１および第２の態様を組み合わせ、第１のフレームと第２のフレームとの対応付けを行った後、単位素片の素片長が音韻継続時間長より小さい場合は第１

２０３に従って選択されたフレームの音源パラメータ２０４と声道パラメータ２０５が合成フレームの合成パラメータとして出力される。

【０００７】音源生成部２３は、フレーム平均ピッチ周期２０１と音源パラメータ２０４に従って、フレーム長２０２の長さの音源信号２０６を出力する。声道フィルタ部２４は、声道パラメータ２０５によって特性が決定される声道フィルタを音源信号２０６で駆動することにより合成音声信号１０９を生成する。

【０００８】フレーム長生成部２６の詳細な動作について説明する。素片情報２０７より得られる当該単位素片の定常部フレーム数および過渡部フレーム数をそれぞれ n_a，n_b とし、音韻継続時間長１１３をＴとすると、合成フレーム長２０２（ｆ）と定常部伸縮率２０８（ｖ）は、次式を満たすように決定される。

【０００９】

$$(v_{na} + n_b)\ f = T \quad (1)$$

のフレームに対応する第２のフレームの単位素片を間引くかまたは繰り返すことにより合成パラメータを生成し、単位素片の素片長が音韻継続時間長以上の場合にはマッピング関数に基づいて第１のフレームのフレーム長を設定し、第１のフレームに対応する第２のフレームの単位素片を合成パラメータとして生成する。

【００１９】このように本発明によると、音声そのものでなく合成パラメータを変化させることで音韻時間長を変化させるため、マッピング関数として例えば傾きが音韻境界付近で予め定められた略一定の値となる関数を用いたり、マッピング関数を単位素片の過渡部と定常部を指定する情報および音韻継続時間長とに従って生成することにより、単位素片の伸縮度合いが連続的に変化することになるため、音韻継続時間長の変化による合成音声の自然性の低下がなく、音質の良好な音声合成が可能となる。

【００２０】

【発明の実施の形態】以下では、音韻（Ｃ，Ｖ）を合成単位とする規則合成器に本発明を適用した場合の実施形態について説明するが、本発明はこのような合成単位に限定されるものではなく、その他の合成単位の場合も適用することが可能である。

【００２１】（第１の実施形態）図１に、本発明の第１の実施形態に係る音声合成器の構成を示す。本実施形態は、合成しようとする音声信号の性質を表す第１のフレーム（合成フレーム）単位の合成パラメータを用いて音声信号を生成する場合、合成フレームのフレーム長を一定とし、合成フレームとなる第２のフレーム（単位素片フレーム）の間引きまたは繰り返しによって、合成音声の音韻継続時間長を変更するようにしたものである。

【００２２】本実施形態において、フレーム平均ピッチ生成部２８、単位素片記憶部２２、音源生成部２３および声道フィルタ部２４の動作は、図９に示した従来の音声合成器と同様である。すなわち、フレーム平均ピッチ生成部２８は、ピッチパターン１１２よりフレーム平均ピッチ周期２０１を求めて出力する。

【００２３】単位素片記憶部２２は、例えば合成単位を音素とすると、全ての母音および全ての子音に対応する単位素片の情報を記憶している。各単位素片は、例えば当該音韻の音声信号を一定のフレーム周期で分析することによって生成され、複数のフレーム（以下、単位素片フレームという）単位で構成される。単位素片フレームは、音源パラメータおよび声道パラメータから構成されており、単位素片記憶部２２はフレーム選択情報２０３に従って選択された単位素片フレームの音源パラメータ２０４と声道パラメータ２０５を出力する。

【００２４】音源生成部２３は、フレーム平均ピッチ周期２０１と音源パラメータ２０４に従って、フレーム長２０２の長さの音源信号２０６を出力する。声道フィルタ部２４は、声道パラメータ２０５によって特性が決定される声道フィルタを音源信号２０６で駆動することにより、合成音声信号１０９を生成する。

【００２５】また、素片情報記憶部６４は、入力された音韻記号列１１１に従って対応する素片の素片情報６０１を出力するが、この素片情報６０１には単位素片フレームのフレーム数、単位素片フレームのフレーム長などの情報が含まれ、定常部の範囲と過渡部の範囲の情報は必ずしも含まれない点で従来の音声合成器と異なっている。

【００２６】次に、本実施形態の特徴的な構成である時間長制御部６０について説明する。時間長制御部６０は、マッピング関数生成部６１、フレーム長生成部６２および合成フレーム選択情報生成部６３から構成されている。マッピング関数生成部６１は、素片情報６０１と音韻継続時間長１１３に従って、合成フレーム列に対応する時間の変数から単位素片のフレームに対応する時間の変数へのマッピング関数６０２を生成する。フレーム長生成部６２は、音韻継続時間長１１３と素片情報６０１に従って、合成フレームのフレーム数６０３およびフレーム長２０２を求める。合成フレーム選択情報生成部６３は、フレーム数がフレーム数６０３で表される合成フレームと、フレーム数がフレーム情報６０１で表される単位素片フレームをマッピング関数６０２を用いて対応付けることによって合成フレーム列を生成し、各合成フレームのフレーム選択情報２０３を出力する。

【００２７】以下、各構成要素の詳細な説明に当たり、記号を次のように定義する。

単位素片フレームのフレーム数：ｎ′
合成フレームのフレーム数　　：ｎ
単位素片フレームのフレーム長：ｆ′
合成フレームのフレーム長　　：ｆ
単位素片長　　：Ｔ′＝ｎ′ｆ′
音韻継続時間長　　：Ｔ
合成フレーム　　：Ｆｊ，（ｉ＝１，２，３，…，ｎ）
単位素片フレーム　　：Ｆ′ｊ，（ｊ＝１，２，３，…，ｎ′）

フレーム長生成部６２は、合成フレームのフレーム数ｎおよびフレーム長ｆを次式に従って求める。

【００２８】

【数１】

$$n = \left[\frac{T}{f'}\right] \quad (2)$$

$$f = T/n \quad (3)$$

【００２９】ただし、記号［］は整数に丸める演算を表すものとする。マッピング関数生成部６１は、合成フレーム列に対応する時間の変数ｔから単位素片のフレームに対応する時間の変数ｔ′へのマッピング関数ｔ′＝ｍ（ｔ）を生成する。マッピング関数ｍ（ｔ）は、定義域が０＜ｔ＜Ｔ、値域が０＜ｍ（ｔ）＜Ｔ′の関数である。また、マッピング関数ｍ（ｔ）は以下の条件式を満たさなければならない。

【００３０】

【数２】

$$m(0) = 0 \quad (4)$$

$$m(T) = T' \quad (5)$$

$$\frac{d}{dt} m(t) = m'(t) \geq 0 \quad (0 \leq t \leq T) \quad (6)$$

【００３１】式（６）は、関数ｍ（ｔ）が単調増加関数であることを表している。ｍ（ｔ）の導関数ｍ′（ｔ）は、マッピング関数ｍ（ｔ）の傾きに対応するため、単位素片の伸縮の度合を表している。すなわち、ｍ′（ｔ）＞１の区間では単位素片のフレームが間引かれ、ｍ′（ｔ）＜１の区間では単位素片のフレームが繰り返されれる。式（４）（５）（６）の条件を満たす関数の例を以下に示す。

【００３２】

【数３】

$$m_1(t) = \begin{cases} \dfrac{2(T-T')}{T^3}t^3 - \dfrac{3(T-T')}{T^2}t^2 + t & (3T' \geq T) \\ \dfrac{4}{27T'^2}t^3 - \dfrac{2}{3T'^2}t^2 + t & (3T' < T,\ t \leq \dfrac{3T'}{2}) \\ \dfrac{T'}{2} & (3T' < T,\ \dfrac{3T'}{2} < t < T - \dfrac{3T'}{2}) \\ \dfrac{4}{27T'^2}(t-3T')^3 - \dfrac{2}{3T'^2}(t-3T')^2 + (t-3T') & (3T' < T,\ T - \dfrac{3T'}{2} \leq t) \end{cases} \quad (7)$$

【００３３】ｍ１（ｔ）の導関数ｍ′１（ｔ）は、０≦ｔ≦Ｔにおいて常に０以上となり、式（６）を満たして

$$m'_1(0) = 1 \quad (8)$$

$$m'_1(T) = 1 \quad (9)$$

いる。また、ｍ′１（ｔ）には次のような特徴がある。

この式は、音韻の境界では単位素片を伸縮させないことを表している。すなわち、式（７）のマッピング関数ｍ１（ｔ）は、音韻の境界付近では単位素片の伸縮の度合が小さく、音韻の中央部では伸縮の度合が大きく、かつ伸縮の度合が連続的に変化するような関数である。

【００３４】合成フレーム選択情報生成部６３は、合成フレームＦｉ，（ｉ＝１，２，３，…，ｎ）と単位素片フレームＦ′ｊ，（ｊ＝１，２，３，…，ｎ′）をマッピング関数ｍ（ｔ）を用いて対応付ける。合成フレームと単位素片フレームの対応付けは、各合成フレームのフレーム中心時刻をｍ（ｔ）によってマッピングし、その時刻が含まれる単位素片フレームとそれぞれ対応付けることによって行う。

【００３５】

【数４】

$$F_i = F'_{ji} \quad (10)$$

$$j_i = \left\lceil \frac{m(f(i-0.5))}{f'} \right\rceil \quad (11)$$

ただし、記号 ⌈ ⌉ は整数に切り上げる演算を表す

【００３６】次に、本実施形態による合成フレーム列の生成例を以下のパラメータの場合について、図２を用いて説明する。

単位素片フレーム数：$n' = 12$

単位素片フレーム長：$f' = 10msec$

単位素片長　　：$T' = n' f' = 120msec$

音韻継続時間長　　：$T = 45msec$

式（７）に$T = 45$，$T' = 120$を代入することによって、マッピング関数$m'_1(t)$は以下のように求められる。

【００３７】

【数５】

$$m_1(t) = -\frac{2}{1215}t^3 + \frac{1}{9}t^2 + t \quad (12)$$

次に、合成フレーム数および合成フレーム長は、式（２）（３）より以下のように求められる。

【００３８】

【数６】

$$n = \left[\frac{45}{10}\right] = 5 \quad (13)$$

$$f = 45/5 = 9 \quad (14)$$

式（１１）を用いてフレームをマッピングする。

【００３９】

【数７】

$$\lceil m_1(9(1-0.5))/10 \rceil = 1 \quad (15)$$

$$\lceil m_1(9(2-0.5))/10 \rceil = 3 \quad (16)$$

$$\lceil m_1(9(3-0.5))/10 \rceil = 6 \quad (17)$$

$$\lceil m_1(9(4-0.5))/10 \rceil = 10 \quad (18)$$

$$\lceil m_1(9(5-0.5))/10 \rceil = 12 \quad (19)$$

【００４０】以上の結果から、$F1 = F'1$，$F2 = F'3$，$F3 = F'6$，$F4 = F'10$，$F5 = F'12$となる。また、本実施形態の変形として、音韻の種類によって異なる種類のマッピング関数を用いることも可能である。例えば、閉鎖区間（無音区間）に続く破裂区間（有音区間）によって構成される無声破裂音などの場合、なるべく閉鎖区間で伸縮させることが望ましいため、後ろほど伸縮の度合が小さくなるようなマッピング関数を用いればよい。

【００４１】（第２の実施形態）図３に、本発明の第２の実施形態に係る音声合成器の構成を示す。本実施形態は、単位素片フレームの間引き／繰り返しは行わず、合成フレームのフレーム長を変化させることによって、合成音声の継続時間長を変化させるようにしたものである。すなわち、合成フレーム数ｎは単位素片フレームのフレーム数ｎ′と等しく、単位素片フレームはそのまま合成フレーム列に対応する。

【００４２】本実施形態において、フレーム平均ピッチ生成部２８、単位素片記憶部２２、音源生成部２３、声道フィルタ部２４および素片情報記憶部６４の動作は、図１に示した第１の実施形態の場合と同様であるため、同一の参照符号を付して説明を省略する。

【００４３】本実施形態の特徴的な構成である時間長制御部７０について説明する。まず、マッピング関数生成部７１は、素片情報６０１と音韻継続時間長１１３に従って、単位素片のフレームに対応する時間の変数から合成フレーム列に対応する時間の変数へのマッピング関数７０１を生成する。フレーム長生成部７２は、マッピング関数７０１と素片情報６０１に従って、合成フレームのフレーム長２０２を求める。合成フレーム選択情報生成部７３は、単位素片フレームのフレーム選択情報２０３を順次出力する。

【００４４】以下、各構成要素の詳細な説明に当たり、記号を次のように定義する。

単位素片フレームのフレーム数：ｎ′
合成フレームのフレーム数　　：ｎ
単位素片フレームのフレーム長：ｆ′
合成フレームのフレーム長　　：ｆｊ，（ｉ＝１，２，３，…，ｎ）
単位素片長　　：Ｔ′＝ｎ′ｆ′
音韻継続時間長　　：Ｔ
合成フレーム　　：Ｆｊ，（ｉ＝１，２，３，…，ｎ）
単位素片のフレーム　　：Ｆ′ｊ，（ｊ＝１，２，３，…，ｎ）

ただし、本実施形態においては単位素片フレームの間引き／繰り返しを行わないため、次式の関係が成立する。

$$n = n' \tag{20}$$

$$F_j = F'_j \quad (i = 1, 2, \cdots, n) \tag{21}$$

マッピング関数生成部７１は、単位素片フレームに対応する時間の変数ｔ′から合成フレーム列に対応する時間の変数ｔへのマッピング関数ｔ＝ｒ（ｔ′）を生成する。マッピング関数ｒ（ｔ′）は、定義域が０＜ｔ′＜Ｔ′，値域が０＜ｒ（ｔ′）＜Ｔの関数である。また、ｒ（ｔ′）は以下の条件式を満たさなければならない。

【００４５】

【数８】

$$r(0) = 0 \tag{22}$$

$$r(T') = T \tag{23}$$

$$\frac{d}{dt} r(t') = r'(t') \geq 0 \quad (0 \leq t' \leq T') \tag{24}$$

【００４６】式（２４）はｒ（ｔ′）が単調増加関数であることを表している。ｒ（ｔ′）の導関数ｒ′（ｔ′）は、マッピング関数ｒ（ｔ′）の傾きに対応するため、単位素片の伸縮の度合を表している。すなわち、ｒ′（ｔ′）＜１の区間では合成フレームのフレーム長が単位素片フレームのフレーム長よりも短くなり、ｒ′（ｔ′）＞１の区間では逆に長くなる。式（２２）（２３）（２４）の条件を満たす関数の例を以下に示す。

【００４７】

【数９】

$$r_1(t') = \begin{cases} \dfrac{2(T'-T)}{T'^3} t'^3 - \dfrac{3(T'-T)}{T'^2} t'^2 + t' & (3T \geq T') \\ \dfrac{4}{27T^2} t'^3 - \dfrac{2}{3T^2} t'^2 + t' & (3T < T',\ t' \leq \dfrac{3T'}{2}) \\ \dfrac{T}{2} & (3T < T',\ \dfrac{3T'}{2} < t' < T' - \dfrac{3T'}{2}) \\ \dfrac{4}{27T^2}(t'-3T)^3 - \dfrac{2}{3T^2}(t'-3T)^2 + (t'-3T) & (3T < T',\ T' - \dfrac{3T'}{2} \leq t') \end{cases} \tag{25}$$

【００４８】ｒ１（ｔ′）の導関数ｒ′１（ｔ′）は、０≦ｔ′≦Ｔにおいて常に０以上となり、式（２４）を満している。また、ｒ′１（ｔ′）には次のような特徴がある。

$$r'_1(0) = 1 \tag{26}$$

$$r'_1(T) = 1 \tag{27}$$

この式は、音韻の境界では単位素片を伸縮させないことを表している。すなわち、式（２５）のマッピング関数は、音韻の境界付近では単位素片の伸縮の度合が小さく、音韻の中央部では伸縮の度合が大きく、かつ伸縮の度合が連続的に変化するような関数である。

【００５０】フレーム長生成部７２は、合成フレームのフレーム長ｆｊ，（ｉ＝１，２，…，ｎ）をマッピング関数ｒ（ｔ′）を用いて生成する。合成フレームのフレ

ーム長ｆｉは、次式より求められる。

$$f_i = r(f'\ i) - r(f'\ (i-1))\ (i=1,\ 2,\ \cdots,\ n) \quad (28)$$

次に、本実施形態による合成フレーム列の生成例を以下のパラメータの場合について、図４を用いて説明する。ただし、図２との比較のため図４は横軸をｔ、縦軸をｔ′として描いている。

【００５２】

単位素片フレームのフレーム数：ｎ′＝５
単位素片フレームのフレーム長：ｆ′＝１０ｍｓｅｃ
単位素片長　　：Ｔ′＝ｎ′ｆ′＝５０ｍｓｅｃ
音韻継続時間長　　：Ｔ＝１２０ｍｓｅｃ

式（２５）にＴ＝１２０，Ｔ′＝５０を代入することにより、マッピング関数は以下のように求められる。

【００５３】

【数１０】

$$r_1(t) = -\frac{7}{6250}t^3 + \frac{21}{250}t^2 + t \quad (29)$$

【００５４】次に、合成フレーム長は式（２８）より以下のように求められる。

$f_1 = 17.28$ [msec]　（３０）
$f_2 = 27.36$ [msec]　（３１）
$f_3 = 30.72$ [msec]　（３２）
$f_4 = 27.36$ [msec]　（３３）
$f_5 = 17.28$ [msec]　（３４）

また、本実施形態の変形として、定常部の範囲と過渡部の範囲の情報を素片情報に含めて、定常部の範囲と過渡部の範囲の情報と音韻継続時間長よりマッピング関数を求めることも可能である。マッピング関数としては、例えば図５で示されるような、過渡部で傾きが常に一定となり、定常部で傾きが滑らかに変化するような関数ｔ＝r_2（ｔ′）を用いることができる。

【００５５】（第３の実施形態）図６に、本発明の第３の実施形態に係る音声合成器の構成を示す。本実施形態は、単位素片の素片長が継続時間長よりも小さい場合は第１の実施形態と同様に動作し、単位素片の素片長が継続時間長以上の場合は第２の実施形態と同様に動作するものである。

【００５６】本実施形態において、フレーム平均ピッチ生成部２８、単位素片記憶部２２、音源生成部２３、声道フィルタ部２４、素片情報記憶部６４、時間長制御部６０および時間長制御部７０の動作は、図１に示した第１の実施形態および図３に示した第２の実施形態の場合と同様であるため、同一の参照符号を付して説明を省略する。

【００５７】モード選択部８１は、素片情報６０１と音韻継続時間長１１３より、単位素片の素片長と音韻継続時間長を比較してモード選択情報８０１を出力する。単位素片の素片長をＴ′とし、音韻継続時間長をＴとすると、このモード選択情報８０１に従って、Ｔ＞Ｔ′の場合には時間長制御部６０が選択され、Ｔ≦Ｔ′の場合には時間長制御部７０が選択される。

【００５８】第１の実施形態では、Ｔ＜Ｔ′の場合にフレーム間引きによる情報欠落の可能性があり、また第２の実施形態では、Ｔ＞Ｔ′の場合にフレーム長が長くなり過ぎてピッチ周期の変化が粗くなる可能性があるが、本実施形態では、第１および第２の実施形態を組み合わせたことにより、情報欠落やピッチ周期の変化が粗くなることがなくなるため、より良好な合成音声が得られる。

【００５９】（第４の実施形態）図７に、本発明の第４の実施形態に係る音声合成器の構成を示す。本実施形態は、第１の実施形態におけるマッピング関数生成部６１をマッピング関数選択部８１およびマッピング関数記憶部８２で置き換えたものである。

【００６０】本実施形態において、フレーム平均ピッチ生成部２８、単位素片記憶部２２、音源生成部２３、声道フィルタ部２４、素片情報記憶部６４、フレーム長生成部６２、合成フレーム選択情報生成部６３の動作は、図１に示した第１の実施形態の場合と同様であるため、同一の参照符号を付して説明を省略する。

【００６１】マッピング関数記憶部８２は、複数のマッピング関数を記憶しており、マッピング関数選択情報８０１に従って選択されたマッピング関数６０２を出力する。マッピング関数選択部８１は、マッピング関数記憶部８２に記憶されているマッピング関数の中から、素片情報６０１と音韻継続時間長１１３に従って、適したマッピング関数を選択し、マッピング関数選択情報８０１を出力する。

【００６２】

【発明の効果】以上説明したように、本発明によれば合成パラメータとなる第２のフレーム単位の単位素片を記憶しておき、入力された音韻継続時間長の情報に従って

生成されたマッピング関数に基づいて第１のフレームと第２のフレームとの対応付けを行い、第１のフレームに対応する第２のフレームの単位素片を合成パラメータとして生成して、この合成パラメータを用いて音声信号を合成することにより、任意の継続時間長の音声を合成することが可能であって、しかも単位素片の伸縮の度合を連続的に変化させることができるため、自然性にすぐれた合成音声を得ることができる。

【図面の簡単な説明】

【図１】本発明の第１の実施形態を示すブロック図

【図２】本発明の第１の実施形態における合成フレーム情報生成の一例を表す模式図

【図３】本発明の第２の実施形態を示すブロック図

【図４】第２の実施形態における合成フレーム情報生成の一例を表す模式図

【図５】第２の実施形態における合成フレーム情報生成の他の例を表す模式図

【図６】本発明の第３の実施形態を示すブロック図

【図７】本発明の第４の実施形態を示すブロック図

【図８】テキスト音声合成システムの構成を表すブロック図

【図９】従来の音声合成器を示すブロック図

【符号の説明】

１…言語処理部

２…韻律処理部

３…合成器

６１，７１…マッピング関数生成部

２２…単位素片記憶部

２３…音源生成部

２４…声道フィルタ部

２５，６４…素片情報記憶部

２６，６２，７２…フレーム長生成部

２７，６３，７３…合成フレーム選択情報生成部

２８…フレーム平均ピッチ生成部

６０，７０，８０…時間長制御部

８１…マッピング関数選択部

８２…マッピング関数記憶部

１０９…合成音声信号

１１１…音韻記号列

１１２…ピッチパターン

１１３…音韻継続時間長

２０１…フレーム平均ピッチ

２０２…合成フレームのフレーム長

２０３…合成フレームのフレーム選択情報

２０４…音源パラメータ

２０５…声道パラメータ

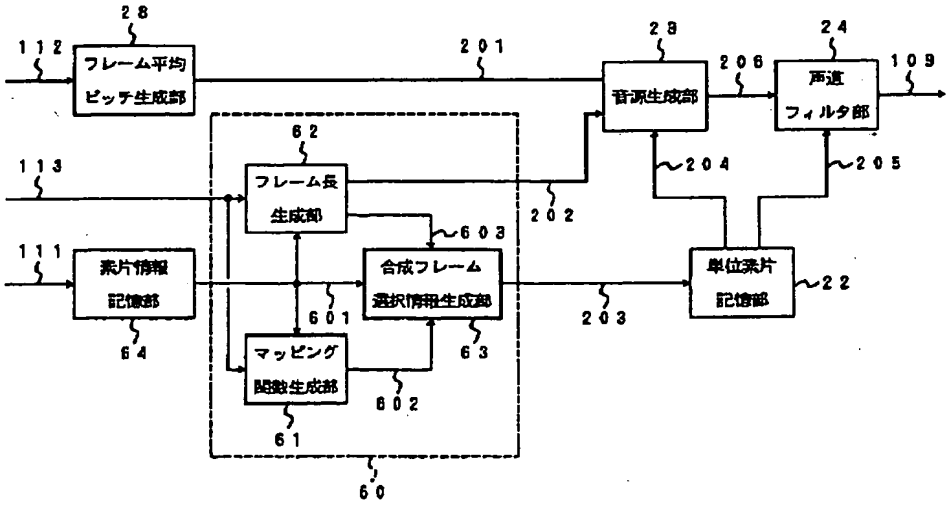
２０６…音源信号

６０１…素片情報

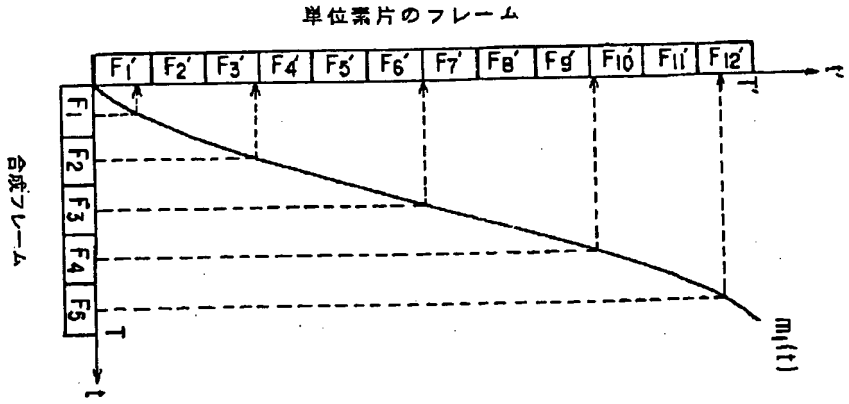
６０２…マッピング関数

６０３…合成フレームのフレーム数

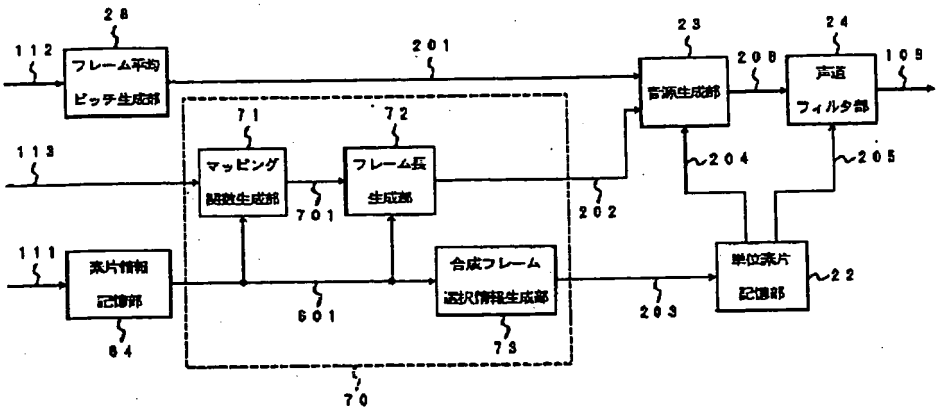
【図１】



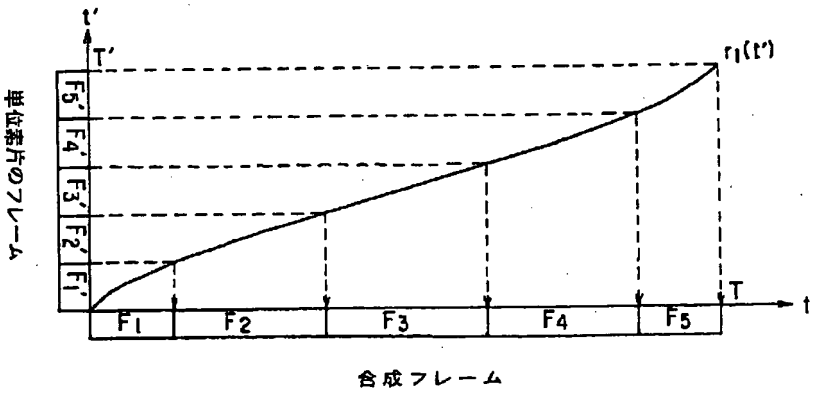
【図２】


単位素片のフレーム
F1' F2' F3' F4' F5' F6' F7' F8' F9' F10' F11' F12'
T'
τ
合成フレーム
F1 F2 F3 F4 F5
T
t
m1(t)

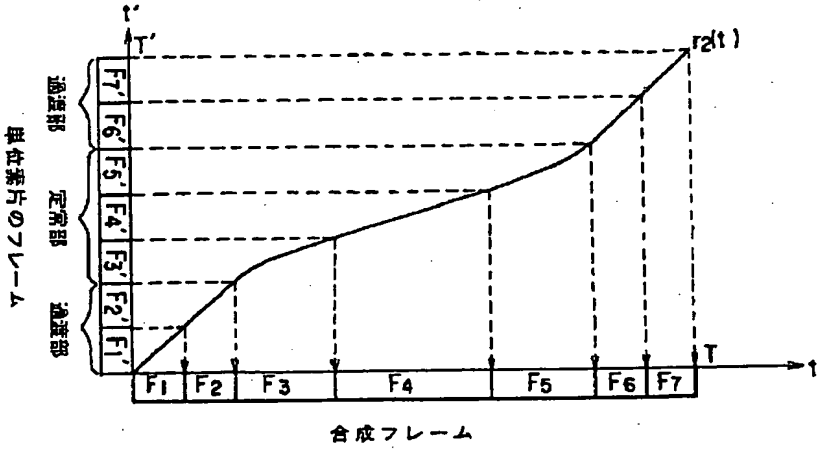
【図３】


28
112
フレーム平均
ピッチ生成部
201
23
音源生成部
208
24
声道
フィルタ部
109
71
72
113
マッピング
関数生成部
フレーム長
生成部
701
202
204
205
111
素片情報
記憶部
601
合成フレーム
選択情報生成部
203
単位素片
記憶部
22
64
73
70

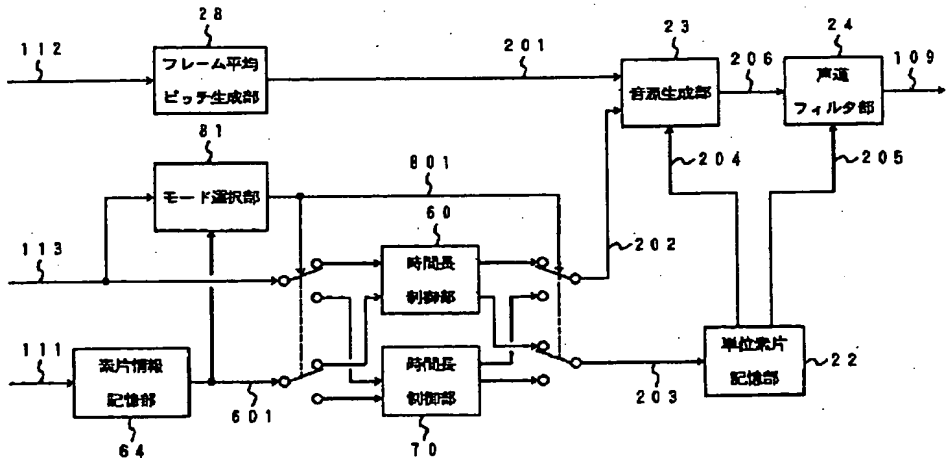
【図４】


t'
T'
r1(t')
単位素片のフレーム
F5' F4' F3' F2' F1'
T
t
F1 F2 F3 F4 F5
合成フレーム

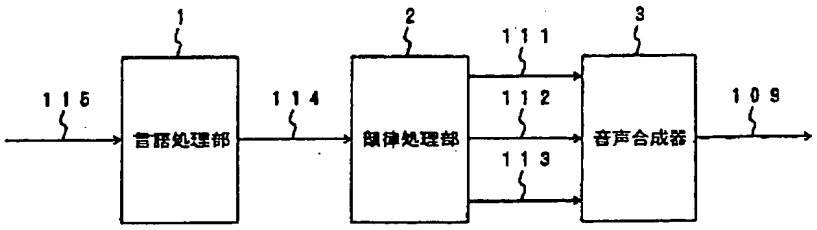
【図５】


t'
T'
r2(t)
過渡部
定常部
過渡部
単位素片のフレーム
F7'
F6'
F5'
F4'
F3'
F2'
F1'
T
t
F1
F2
F3
F4
F5
F6
F7
合成フレーム

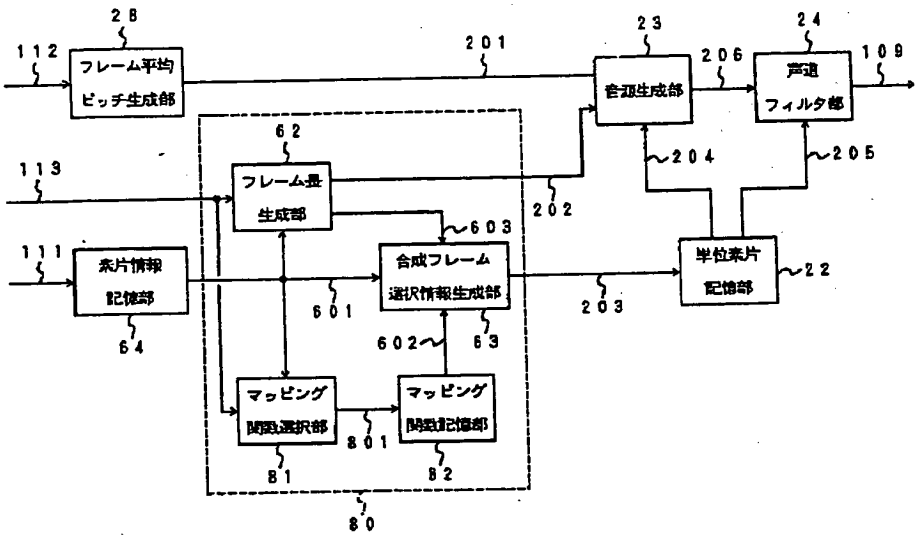
【図６】


28
112
フレーム平均
ピッチ生成部
201
23
音源生成部
206
24
声道
フィルタ部
109
81
モード選択部
801
60
204
205
113
時間長
制御部
202
111
素片情報
記憶部
601
時間長
制御部
203
単位素片
記憶部
22
64
70

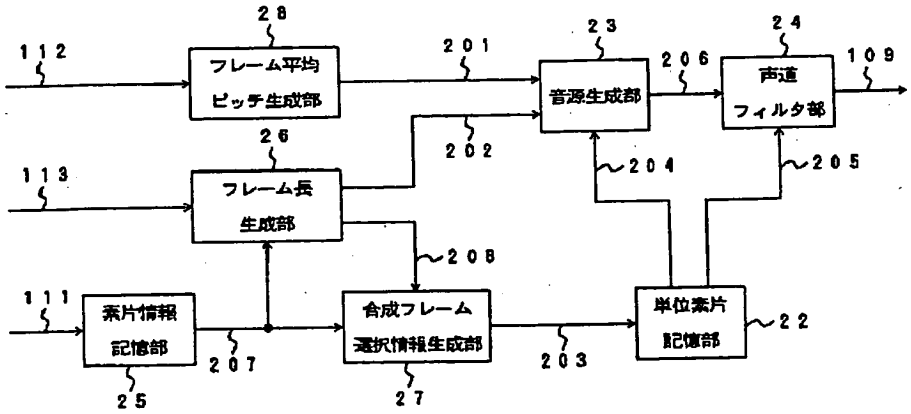
【図８】


1
2
111
3
115
言語処理部
114
韻律処理部
112
音声合成器
109
113

【図７】


28
112
フレーム平均
ピッチ生成部
201
23
音源生成部
206
24
声道
フィルタ部
109
62
113
フレーム長
生成部
603
202
204
205
111
素片情報
記憶部
601
合成フレーム
選択情報生成部
203
単位素片
記憶部
22
64
602
63
マッピング
関数選択部
801
マッピング
関数記憶部
81
82
80

【図９】


28
112
フレーム平均
ピッチ生成部
201
23
音源生成部
206
24
声道
フィルタ部
109
202
26
113
フレーム長
生成部
204
205
208
111
素片情報
記憶部
207
合成フレーム
選択情報生成部
203
単位素片
記憶部
22
25
27

METHOD OF VOICE SYNTHESIS

Patent number: JP10091191
Publication date: 1998-04-10
Inventor: KAGOSHIMA TAKEHIKO; AKAMINE MASAMI
Applicant: TOSHIBA CORP
Classification:
- international: G10L5/04
- european:
Application number: JP19960246718 19960918
Priority number(s):

Abstract of **JP10091191**

PROBLEM TO BE SOLVED: To provide a method of voice synthesis permitting to generate a synthesized voice for an arbitrary phoneme duration without causing damage to the nature. SOLUTION: In a method of a voice synthesis for synthesizing voice signals 10 by using synthesis parameters (sound source parameter 204, voice path parameter 205) of a first frame unit representing the nature of voice signals to be synthesized, synthesis parameters are generated by storing unit element pieces of a second frame unit to become synthesis parameters in a unit element piece storage part 22 beforehand, giving a synthesis frame selection information generation part 63 a mapping function generated by a function generation part 61 according to information of the inputted phoneme duration 113, making the first and second frames corresponded to each other, and further, thinning out or repeating the unit element pieces of the second frame corresponding to the first frame according to the mapping function.